UNIVERSITY OF ILLINOIS

Agricultural Experiment Station

BULLETIN No. 148

ON THE MEASUREMENT OF CORRELATION WITH SPECIAL REFERENCE TO SOME CHARACTERS OF INDIAN CORN

By HENRY L. RIETZ and LOUIE H. SMITH

URBANA, ILLINOIS, NOVEMBER, 1910

CONTENTS OF BULLETIN NO. 148

ON THE MEASUREMENT OF CORRELATION WITH SPECIAL REFERENCE TO SOME CHARACTERS OF INDIAN CORN

By Henry L. Rietz, Statistican, and Louie H. Smith, Assistant Chief, Plant Breeding

Introduction

In Bulletin 119 of this station, there are presented methods of dealing with problems involving variability of a single character and these methods are there applied to the study of type and variability of some characters in corn.

But the breeder deals with many characters in the same organism, each with its own variability. After treating separate characters, what he needs next to know is whether, and to what extent, any bond may exist between characters by virtue of which, if one character varies, other characters of the same organism tend also to move in the same or in opposite directions.

If such a bond exists the characters are said to be correlated (co-related), and it is the purpose of this bulletin to describe methods by which such a correlation may be detected if present and the strength of its bond be measured.

A second purpose of the bulletin is to present data concerning certain definite correlations for corn bred at the Illinois Station.

The great value to the breeder of definite knowledge of correlations within a species is that it gives reliable information, enabling him to predict from the presence of certain characters the most probable values of associated characters.

More technically speaking, when we are dealing with two systems of variable characters in correspondence, we are, in general, much concerned about whether fluctuations of variates in one system are in sympathy with fluctuations of corresponding variates in the other, and with establishing causal relations between the two series of phenomena. For example, in breeding corn for commercial purposes, we are much interested in knowing what characters of the seed ears should be modified or selected to increase the yield; and, if we should select directly one character, it is important to know to what extent other characters are being selected indirectly, because of the tendency of the two to fluctuate in the same or in opposite directions.

These examples illustrate the following technical definition of *correlation: Two characters*—say length and circumference of ears of corn—*are said to be correlated when with any selected values (x) of the one character, we find that values of the other character, a given amount above and below the mean of that character, are not equally likely to be associated.*

As the first and simplest method, it may possibly occur that correlation is so pronounced that it may be necessary merely to look at two sets of figures to note that corresponding values have a tendency to change simultaneously in the same or in opposite directions. The existence of such decided correlation may be known by inspection.

As a second, and somewhat more effective method, one may plot curves for each of two systems of variates, and if correlation is very pronounced, it may sometimes be discovered by noting whether the curves have a tendency to rise and fall together, or if, when one rises, the other falls.

Not only do these methods prove inadequate to detect correlation unless it is exceedingly pronounced, but they lack precision in that they do not give a measure of correlation. It is not enough to know whether correlation exists, its quantitative measure is usually a matter of importance.

Our power to measure the correlations among associated phenomena has been enormously increased during the past two decades by methods introduced by Galton and developed by Pearson and those associated with him. An application of these methods has not until very recently been made to problems in agriculture.*

It is the purpose of this Bulletin to present in a form useful to agricultural students the methods of correlation measurement without presuming more mathematics than is absolutely necessary, and to give the results of our investigations into the correlation of certain characters in corn bred at the Illinois Station.

1. The Correlation Table.—The first step in the process of measuring correlation is to construct a double entry table (Fig. 1) —called a "correlation table"—out of the measurements of the characters in a large number of individuals. One mark at the intersection of the proper column and row in the table records a pair of corresponding variates with reference to two characters.

Put in tabular form as it appears in the actual work, we have the following (Fig. 1) for the correlation table between the num-

*Davenport, *Principles of Breeding*, pp. 452-472, 703-711.
Pearl and Surface, *Bulletin 166, Maine Agr. Exp. Sta.*
Clark, *Bulletin 279, Cornell University Agr. Exp. Sta.*

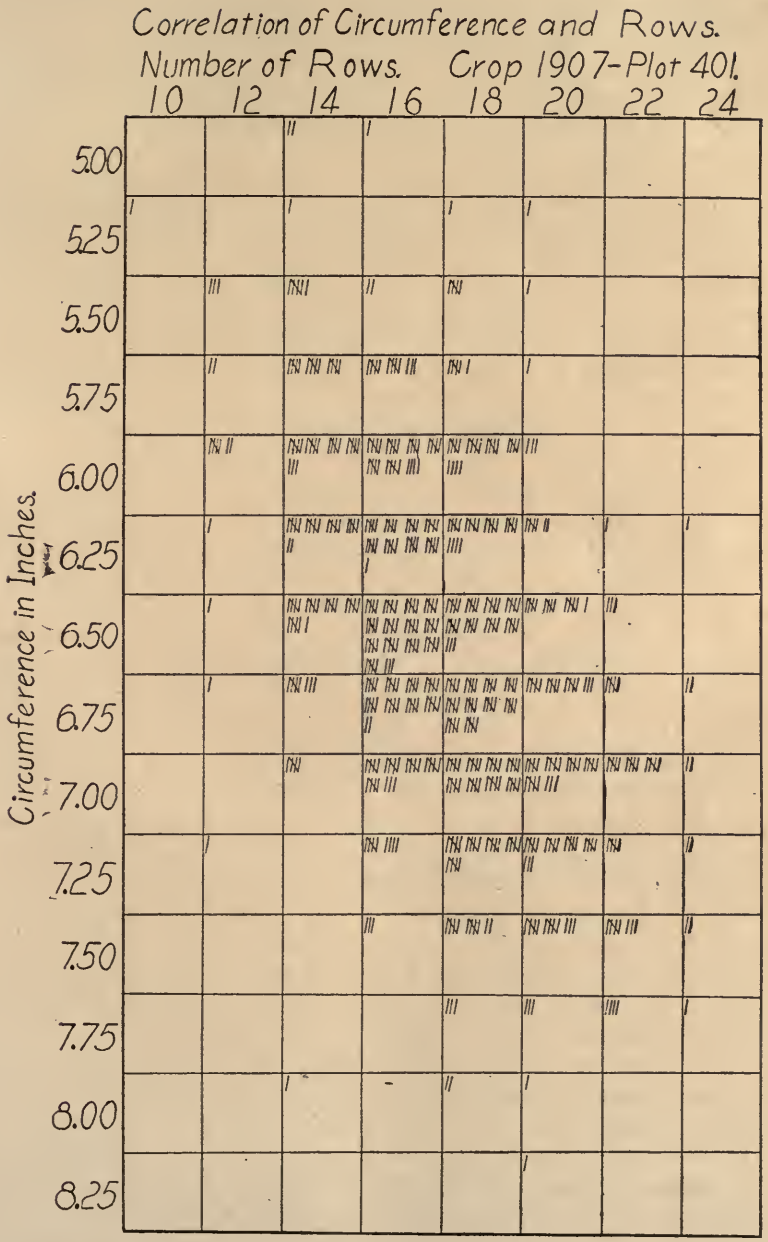
Correlation of Circumference and Rows.
Number of Rows. Crop 1907-Plot 401.
10
12
14
16
18
20
22
24
Circumference in Inches.
5.00
5.25
5.50
5.75
6.00
6.25
6.50
6.75
7.00
7.25
7.50
7.75
8.00
8.25

FIG. 1.

ber of rows of kernels on ears and the circumference of the ears for a certain plat of corn (Plot 401) grown in 1907 at the Illinois Experiment Station.

It will be observed that this table consists of a double system of arrays each of which is a frequency distribution as explained in Bul. 119, and has its own mean and standard deviation as has any other frequency distribution.

To show, in a concrete way, how such a table is made, suppose an ear of corn has 18 rows of kernels and a circumference between 6.375 and 6.625 in., a mark is made in the rectangle at the intersection of the column headed 18 and the row of the table marked 6.50. A second table (Fig. 2) exhibits the result of counting the marks in each of the rectangles of Fig. 1.

Any number in this table, say 43, in the column headed 18 and the row marked 6.50, indicates that 43 ears of the total of 769 ears had 18 rows of kernels and a circumference of class mark 6.50.

By adding the numbers in horizontal arrays, we obtain the frequency distribution of the population with respect to circumference of ears, and by adding the numbers in columns, we obtain the frequency distribution of the population with respect to rows of kernels on ears (See Fig. 3).

The mere superficial inspection of a correlation table may suggest that a certain amount of correlation exists. For example, ears of corn of circumference 7.50 inches, from this population, are much more likely to have 20 rows of kernels than are ears of circumference 6 inches. It is pretty clear that there is a tendency, in general, for the marks in the table of Fig. 1 to arrange themselves in a region along the diagonal from the upper left hand corner to the lower right hand corner of the table. This signifies that a positive correlation exists; that is, in general, for this population, ears that have a large number of rows of kernels are more likely to be large in circumference than are ears with a smaller number of rows of kernels. But it is not our purpose merely to detect the existence of correlation. What we seek is a statistical *coefficient* that will serve to measure correlation, and that will enable us to predict with as high a degree of probability as possible, from an assigned character, the value of the associated character in the related system of variates. The coefficient of correlation, denoted by r in this paper, is useful for this purpose.

2. Nature of the coefficient r.—A discussion of the mathematical theory of correlation will be given in the Appendix to this Bulletin, but the general character and common sense significance of r may well be stated here. The value of the coefficient is within the limits —1 and +1. If r=1, there is said to be perfect positive

Correlation of Circumference and Rows.
Number of Rows. Crop 1907–Plot 401.

Circumference in Inches.	10	12	14	16	18	20	22	24
5.00			2	1				
5.25	1		1		1	1		
5.50		3	6	2	5	1		
5.75		2	15	13	6	1		
6.00		7	23	34	24	3		
6.25		1	22	41	24	7	1	1
6.50		1	26	68	43	16	3	
6.75		1	8	42	50	18	5	2
7.00			5	28	40	28	15	2
7.25		1		9	25	23	5	2
7.50				3	12	13	8	2
7.75					3	3	4	1
8.00			1		2	1		
8.25						1		

FIG. 2.

correlation; that is, for any assigned value of the character in one system, the value for each corresponding individual of the related system is known, and the ratio of the deviations of any two variates of a pair from their mean values is a constant for all pairs. In other words, perfect correlation (r=1) indicates complete causation in the sense that the two characters go together perfectly. If r=—1, there is said to be perfect negative correlation. In this case, the ratio of the deviations from mean values are *negative* and constant for all pairs. If no correlation exists, the two characters appear indifferent to each other, and this fact is expressed by r=0. In a general way, we may say that the correlation should be judged, in any application, by the value that r takes between —1 and +1. For our applications to characters in corn, there is usually a positive correlation, and the amount of correlation is measured by the value of r between 0 and 1.

The correlation coefficient may be defined as the mean product of deviations of corresponding variates from their mean values in units of the standard deviations.

The meaning of the standard deviation of a frequency distribution is shown in Bulletin 119 of this Station. If a variate is below the mean, its deviation is negative; while if it is above the mean, it is positive. Hence, if each individual of a pair of variates is above the mean of the system to which it belongs, or if each of the pair is below, the pair tends to contribute to positive correlation. On the other hand, if one variate of a pair is below the mean of its system and the other above, the product is negative, and such a pair tends to contribute to negative correlation. While it appears from this that the coefficient of correlation, as defined above, has a common sense justification, we shall require the mathematical methods of the Appendix to see more fully how this coefficient with the standard deviations of the two systems of variates are descriptive of the correlated population exhibited on a correlation table such as is shown in Fig. 1.

3. Details of the computation of r.—In algebraic form

$$r = \frac{\Sigma xy}{n\sigma_x \sigma_y} \quad \text{----(1)}$$

where Σxy means the sum of the products of the deviations of corresponding variates from their mean values; and σ_x, σ_y are standard deviations, while n is the number of pairs of variates.

As we shall use the correlation table of Fig. 2 to illustrate a systematic arrangement of the work in the computation of r, the formula (1) may appear more significant in the application by writing it in the form

$$r = \frac{\Sigma D_c D_R}{n\sigma_c\sigma_R} \text{ - - - - (2)}$$

where the subscripts c and R refer to circumference and rows of kernels respectively, while the D's represent deviations of characters indicated by subscripts. That is to say, D_c is the deviation of the circumference of an ear from the mean circumference, and D_R the deviation of the number of rows of kernels from the mean of the number of rows of kernels.

There is derived in the Appendix, pp. 313-314, a formula which gives the same numerical value as

$$\frac{\Sigma xy}{n\sigma_x\sigma_y} \quad \text{or} \quad \frac{\Sigma D_c D_R}{n\sigma_c\sigma_R},$$

and, while its algebraic expression is a little more complicated, it is much better adapted to numerical computation than the above formula, as it avoids the use of decimals until almost the end of the work. In this respect, it is analogous to the shorter method presented in Bulletin 119 for finding the mean and the standard deviation. If applied to the case of the number of rows of kernels and circumference of ears of corn, the formula is

$$r = \frac{1}{\sigma_c\sigma_R}\left(\frac{\Sigma D_R' D_c'}{n} - C_R C_c\right) \text{ - - - - (3)}$$

where D_R', D_c' are deviations from our guesses at the means instead of deviations from the means themselves; and C_R, C_c are the corrections applied to the guesses at the mean number of rows of kernels and circumference respectively in finding the means and standard deviations.

In the actual work of calculating r from formula (3), it is highly important to have a systematic form in which to arrange the work, in order to avoid confusion in the somewhat complicated details. It seems desirable, for this reason, to describe the arrangement of the actual work as shown in Fig. 3.

Having given the correlation table, we first add the numbers in the arrays with respect to both characters; that is, add numbers in rows and columns of the table. This gives two frequency distributions—the one with respect to circumference exhibited in the vertical column headed f_c, and the other with respect to rows of kernels shown in the horizontal column of figures marked f_R. For each of these frequency distributions, the *means and standard deviations* are calculated by the shorter method explained and applied in Bulletin 119.

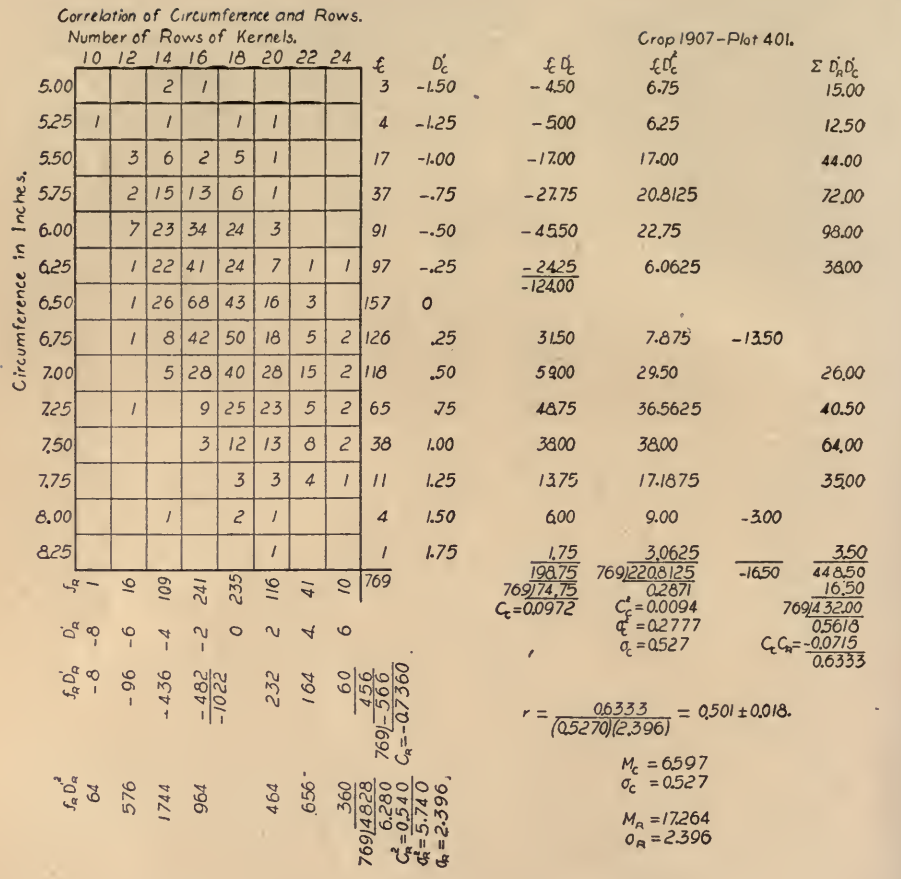

FIG. 3.

The results are

$$M_c = 6.597,$$
$$\sigma_c = 0.527,$$
$$M_R = 17.264,$$
$$\sigma_R = 2.396,$$

where M_c, M_R represent mean circumference and mean number of rows of kernels respectively, while σ_c, σ_R are corresponding standard deviations.

The columns of figures marked f_c, D_c', f_cD_c', $f_cD_c'^2$, f_R, D_R', f_RD_R', $f_RD_R'^2$ are all self explanatory to one familiar with the meaning of algebraic symbols, and who knows how to find the mean and standard deviation.

There remains the column of figures headed $\Sigma D'_R D'_c$, which we shall endeavor to explain in detail, as this is the only part of the computation that is actually new to one who knows how to calculate the variability of a population. In finding the means, we get the deviations of class marks from our guesses at mean circumference and mean number of rows of kernels on ears of corn. These deviations are marked D'_c, D'_R. For example, in row 1, we find 3 ears of circumference 5 inches. These three ears deviate —1.50 from our guesses at the mean circumference.

We next form the product of each number of the correlation table and of the two corresponding deviations. For example, where the column is headed 16 and the row is labelled 6.00 inches intersect, occurs the number 34. These 34 ears have deviations from our guesses of —0.50 and —2 as is indicated by the symbols D'_c and D'. Hence, for this number 34, we form the product $34(-0.50)(-2)=+34$. Without regard to labor, we should find such a product for each compartment of the correlation table. The sum of these products, with due regard to signs, is the $\Sigma D'_R D'_c$ of formula (3). The systematic way to carry out this work is to record the results of this operation for each horizontal array in line with the array under the heading $\Sigma D'_R D'_c$, and then to add the results for separate arrays to obtain 432.00 which is symbolically ndicated by $\Sigma D'_R D'_c$.

To illustrate the method of calculation, let us take the array of circumference 6.00 inches as an example. We have, for this array,

$$\left.\begin{array}{l} -6\times 7 \\ -4\times 23 \\ -2\times 34 \\ 0\times 24 \\ 2\times 3 \end{array}\right\}\times(-0.50)$$

This gives 98.00,

For the array of mark 6.75

$$\left.\begin{array}{l} -6\times 1 \\ -4\times 8 \\ -2\times 42 \\ 0\times 50 \\ 2\times 18 \\ 4\times 5 \\ 6\times 2 \end{array}\right\}\times(0.25)$$

This gives —13.50.

Treat all arrays in this manner, and divide the sum of the products thus obtained (that is, 432.00) by the number of variates 769. This gives $\frac{\Sigma D_c' D_R'}{n}$ of formula (6) and equals 0.5618.

Next, we subtract from this the product of our two corrections in finding means. That is, $C_R C_c = -0.0715$.
To subtract this negative number, we must add 0.0715. This gives 0.6333 for the numerical value of $\frac{\Sigma D_c' D_R'}{n} - C_c C_R$.

$$r = \frac{1}{\sigma_R \sigma_c}\left(\frac{\Sigma D_c' D_R'}{n} - C_c C_R\right) = 0.501.$$

Correlation of Circumference and Rows. Crop 1907 – Plot 401.

Circumference in Inches	Number of Rows of Kernels. 10	12	14	16	18	20	22	24	f_c	D_c'	$f_c D_c'$	$f_c D_c'^2$	$\Sigma D_R' D_c'$
5.00			2	1					3	-6	-18	108	30
5.25	1		1		1	1			4	-5	-20	100	25
5.50		3	6	2	5	1			17	-4	-68	272	88
5.75		2	15	13	6	1			37	-3	-111	333	144
6.00		7	23	34	24	3			91	-2	-182	364	196
6.25		1	22	41	24	7	1	1	97	-1	97	97	76
6.50		1	26	68	43	16	3		157	0	-496		
6.75		1	8	42	50	18	5	2	126	1	126	126	-27
7.00			5	28	40	28	15	2	118	2	236	472	52
7.25		1		9	25	23	5	2	65	3	195	585	81
7.50				3	12	13	8	2	38	4	152	608	128
7.75					3	3	4	1	11	5	55	275	70
8.00			1		2	1			4	6	24	144	-6
8.25						1			1	7	7	49	7
											795	3533	-33
											769|299	769|4594	897
											$C_c = 0.389$	$C_c^2 = 0.151$	33
												$\sigma_c^2 = 4.443$	769|864
												$\sigma_c' = 2.108$	1.1235
													-0.1430
													$C_c' C_R = 1.2665$

	10	12	14	16	18	20	22	24	
f_R	1	16	109	241	235	116	41	10	769
D_R'	-4	-3	-2	-1	0	1	2	3	
$f_R D_R'$	-4	-48	-218	-241 / -511		116	82	30 / 228	769|-283
									$C_R' = -0.3680$
$f_R D_R'^2$	16	144	436	241		116	164	90	769|1207
									1.570
									0.135
									$\sigma_R^2 = 1.435$
									$\sigma_R' = 1.198$

$$r = \frac{1.2665}{(2.108)(1.198)} = 0.501 \pm 0.018$$

Fig. 4

4. Modification of the method of computing r (Fig. 4).—The computation of r may, in general, be further simplified by taking the difference between two successive class marks as the unit of measurement. That is, the units of grouping are the units throughout. Then, as shown in Fig. 4, the deviations from the guesses are consecutive integers. This method gives precisely the same value for the correlation coefficient as that explained under Fig. 3. But the standard deviations and the corrections to guesses at the means are expressed in terms of differences between consecutive classes as units. When thus expressed, we note on Fig. 4, that

$$\sigma_c' = 2.108,$$
$$\sigma_R' = 1.198.$$

Where σ_c' is the standard deviation in circumference (expressed in units equal to the difference between classes) and σ_R' is the standard deviation in rows of kernels (similarly expressed). To express σ_c' in inches, we must multiply by 0.25. This gives $\sigma_c = 0.527$ as before.

Similarly, to make σ_R' consistent with σ_R of figure 3, we must multiply by 2. This gives

$$\sigma_R = 2.396.$$

5. Probable error.—It still remains to find the probable error in our computed value. The general meaning of the probable error, and its use in indicating the degree of confidence to be placed in a result obtained from a random sample of a population has been given in Bulletin 119. It seems sufficient here to give merely the formula for the probable error in the coefficient of correlation r. This formula is

$$E_r = \pm \frac{0.6745\,(1-r^2)}{\sqrt{n}}$$

Applied to our example, $E_r = \dfrac{0.6745\,[\,1 - (0.501)^2\,]}{\sqrt{769}}$

$$= 0.018$$

Hence, we write

$$r = 0.501 \pm 0.018$$

as the measure of the correlation in question.

6. Use of the correlation coefficient.—The general use of such a precise measure of correlation as is given by r, has perhaps been sufficiently discussed in the introduction. However, it seems well to emphasize here that in the selection of one character, we, in general, indirectly select correlated characters, and change their means and standard deviations accordingly. Again, if in

the selection of what the breeder conceives to be the most desirable type for parents, he artificially increases or decreases correlations between characters, this process will, in general, change the correlation between these characters in parents and in offspring. For example, it appears that the correlation between length of ears and circumference for different types of corn which we have examined varies between the limits 0.128 and 0.623. Now, we have examined cases in connection with this work where 48 parents are selected so as to exhibit a negative correlation of —0.21 between length and circumference. If we use the offspring of such a set of 48 ears and ask to what extent length and circumference are inherited, or if we ask to what extent these characters in the parent are correlated with the yield, it is pretty clear that we have much complicated our problem by the selection and imposition of the correlation —0.21. Hence, it is highly desirable to know what correlations actually exist among different characters, before we should expect to obtain in even an approximately precise way the correlation between parent and offspring (inheritance) or the correlation between characters in the parent and yield.

7. The regression coefficient.—From the correlation coefficient and the standard deviations of each of the two characters, it is easy to obtain what is known as the *regression* coefficient. For example, to obtain the regression coefficient of circumference relative to the number of rows of kernels, multiply the coefficient of correlation by the standard deviation in circumference and divide the product by the standard deviation in the number of rows of kernels. This gives, for the particular example above,

$$r \frac{\sigma_C}{\sigma_R} = 0.110.$$

Similarly, the regression of the number of rows of kernels relative to the circumference of ears is

$$r \frac{\sigma_R}{\sigma_C} = 1.200.$$

8. Use of the regression coefficient.—In many systems of correlated variates, the regression is of a kind described in the appendix as *linear regression.* In such cases, the regression coefficient gives us a useful method of predicting, from a given value of one character, the most probable value of the corresponding correlated character. That is to say, from the selected value of one character, we calculate the mean value of the corresponding array.

For the particular case in hand, suppose we select ears with twenty rows of kernels, such ears deviate 2.736 above the mean number of kernels on ears, the regression coefficient (0.110) of circumference on number of rows of kernels indicates that we should expect the mean circumference of ears of 20 rows of kernels to be (0.110) (2.736)=0.301 inches above the mean circumference for the entire population. That is, ears with 20 rows should have a mean circumference

$$6.597+0.301=6.898.$$

By actual computation, the mean of the array of class mark 20 rows is 6.927 ± 0.121, so that the regression coefficient gives the mean of the array to within deviations due to random sampling.

To be more general, if we select an ear of any deviation x in the number of rows of kernels from the mean, we should expect the deviations in circumference of corresponding ears to center about 0.110x.

Similarly, if we select an ear of any deviation y in circumference from the mean circumference, we should expect the deviation in number of rows of kernels to center about

$$1.20\ y.$$

In beginning the discussion of the use of the regression coefficient, we limited our remarks to linear regression. This means that if the correlation table is constructed to scale, and the mean values of arrays be plotted, these mean values will lie along a straight line to within deviations to be attributed to random sampling. Fortunately, this condition is, in general, well satisfied in our applications.

It is important in every case to examine the correlation table to ascertain whether the means of systems of parallel arrays lie reasonably near a straight line.

9. Determination of the correlation coefficients for certain physical characters in corn.—Corn grown on experimental plots of the Illinois Station in 1907, 1908, 1909, furnishes the material for the present study of the correlation between physical characters in corn, and for the problem of quantitative laws of inheritance of these characters. It should be understood that the problem of inheritance is a problem of the correlations between ancestry and offspring. The characters with which we propose to deal here are: length, weight, circumference of ears, and the number of rows of kernels on ears.

While we have done considerable work on the inheritance of these characters and hope to publish these results, it appears better to present in the present bulletin the correlations between characters as we find them in large populations, with very little reference to heredity. We do this because, with our material, as is very

general in the quantitative study of inheritance in plants, the question of the precision of the results is complicated by the fewness of parents relative to the number of offspring, as well as by the rather stringent selection of parents. We think it expedient to defer to a later bulletin the treatment of these difficulties. Further, it is important to know these correlations before attempting a study of inheritance or of the correlation between characters in parent ears and yield.

In the comparison of two statistical results, the difference between the two results compared to its probable error is of great value. In general, we may take the probable error in a difference to be the square root of the sum of the squares of the probable errors of the two results. If the difference does not exceed two or three times the probable error thus obtained, the difference may reasonably be attributed to random sampling. If the difference between the two results is as much as 5 to 10 times the probable error, the probabilty of such differences in random sampling is so small that we are justified in saying that the difference is significant. In fact, a difference of ten times its probable error is certainly significant in so far as there is certainty in human affairs.

Such significant differences in our applications may perhaps be well divided into three classes:

(1) Those due to differences in variety of corn.
(2) Those due to seasonal influences.
(3) Those due to difference in soil treatment.
(4) Those to be attributed to selection of parents.

10. Source of material.—The material for this study is furnished by the crops obtained from a number of the regular experiment plots which are being conducted for different purposes. These plots may be considered as belonging to two different groups, one of which is devoted primarily to soil investigation and the other to experiments in corn breeding.

The soil plots comprise what are designated as the 400 and 500 series. They are devoted to a two-year rotation consisting of corn alternating with oats, that is to say, corn occupies the 400 series one year and the 500 series the next year.

Each series is divided into ten plots of one-tenth acre numbered from 401 to 410 and from 501 to 510. To these plots various soil treatments have been applied as follows:

401 and 501—None (check plot).
402 " 502—Legume catch-crop and crop residues.*

*Corn stalks and oat straw plowed under, removing only the grain from the land.

403 and 503—Farm manure.
404 " 504—Legume and crop residues and lime.
405 " 505—Manure and lime.
406 " 506—Legume and crop residues, lime and phosphorus.
407 " 507—Manure, lime and phosphorus.
408 " 508—Legume and crop residues, lime, phosphorus and potassium.
409 " 509—Manure, lime, phosphorus and potassium.
410 " 510—Legume and crop residues with extra heavy manure and phosphorus.*

The yields from these variously treated plots are given in the following tables in connection with the other data.

This brief description will serve to explain in a general way the significance of the various plots and the following data pertaining to them. The reader who may be interested in a more detailed account regarding the arrangement, description and history of these soil plots is referred to Bulletin 125 of this Station,—"Thirty Years of Crop Rotations on the Common Prairie Soils of Illinois," where are given the complete records.

The particular variety of corn grown upon these plots has been two strains of Leaming which have been under selection for a number of years for high-protein and low-protein content respectively, the work being controlled by the method of "mechanical selection" described in Bulletins 55-87-100.

To be more definite, in 1907, 1909, seed corn low in protein content was planted, while in 1908, seed corn high in protein was planted on the 400 and 500 series concerned in this investigation.

Although a material difference in composition between the two strains has been effected through this method of selection, this difference does not seem to have significantly affected the correlation values under consideration, as will appear from the results of this bulletin.

The corn breeding plots from which samples have been taken for these correlation studies represent four lines of selection which have been under way since 1896, the object in view being to change the normal composition of the grain of a variety of corn by producing strains of special chemical characteristics.

In this manner four strains of markedly different chemical composition have arisen from a single variety by selecting continuously for

*Five times the ordinary application of manure and of phosphorus.

1—High-protein content
2—Low-protein content
3—High-oil content
4—Low-oil content.

For the history and the results of the first ten generations of this work the reader is referred to Bulletin 128, "Ten Generations of Corn Breeding." The variety under experiment contained originally in 1896 an average of 10.92 percent of protein and 4.70 percent of oil.

The composition of the different strains for the years herein concerned is as follows:

	High Protein.	Low Protein.	High Oil.	Low Oil.
1907	13.89	7.32	7.43	2.59
1908	13.94	8.96	7.19	2.39
1909	13.41	7.65	6.96	2.35

11. Discussion of results.—With a set of the four characters under consideration, there are possible six pairs of variates between each pair of which we can determine the correlation. From the results of the tables, it will be observed that we have carried out the determination of the correlation coefficient for some plots for each of these six pairs of characters. As the correlations coefficients we have determined and given in this paper seem sufficient to give a good general notion as to the value of correlations between these characters, it has appeared as well to defer further calculations of correlation coefficients until we ascertain whether further special determinations will be of service in problems of inheritance of these characters and their correlations with yield,—the problems to which we regard the present investigation as preliminary. The accompanying tables include the results of 141 determinations of correlation.

Two year rotation corn.—In length and circumference, the correlation centers about 0.33 for the year 1907, 0.47 for 1908, and 0.49 for 1909. For the three years together, the values center about 0.43. The smallest correlation is given by data from plot 409 of 1907. This correlation is 0.203. The greatest correlation is 0.623 furnished by data of plot 402 in 1909. These extreme values are very different as seen by comparison with their probable errors. They belong to different years, and the value 0.623 corresponds to a decidedly low yield while 0.203 corresponds to a high yield. In fact, there seems to be a somewhat general tendency towards high correlation of length and circumference when the yield is low and vice versa. There are three small values of correlation between length and circumference. These belong to

three consecutive plots of 1907. Inspection of the correlation table shows that, in these cases, there is considerable deviation from linear regression—the ears of extremely large circumference tended to be shorter than ears of less extreme circumference.

In length and number of rows, the correlations are insignificant except possibly in one case where r is more than four times its probable error.

In circumference and rows, the correlation centers about 0.486 in the year 1907, 0.499 in 1908, 0.467 in 1909. The extremes presented are 0.425 and 0.608, which do not differ so much as the extremes for length and circumference. While the deviations are too great to be assigned to random sampling, it appears that there is both a difference in season, soil, and parentage.

In length and weight of ears, the mean value of the correlation is 0.810 in 1909, and these correlations did not show very great differences for different plots. We may regard 0.8 as sort of rough value for this correlation.

In weight and rows of kernels, we have values from 0.178 to 0.345. In weight and circumference, we have values from 0.648 to 0.840.

The Illinois Corn.—In length and circumference, the correlation are very different for selected strains. The conditions are complicated by differences of soil, and season. The low oil plot of 1907 gives the lowest correlation, whereas the low oil plot of 1908 gives the highest correlation of the entire series.

In correlation between circumferences and rows of kernels, there are much smaller differences between plots than for length and circumference.

Between length and rows of kernels, there appears to be no significant correlation, except possibly in one case.

Between length and weight, the correlations are not very different, and fall roughly between 0.65 and 0.85.

Arranging the pairs of systems of variates in descending order as to correlation, we have the following order:

(1) Length and weight.
(2) Circumference and weight.
(3) Circumference and rows of kernels.
(4) Length and circumference.
(5) Weight and rows of kernels.
(6) Length and rows of kernels.

For this arrangement the odds are pretty large except in the case of (3) and (4), and possibly (1) and (2). As a sort of general conclusion, we may say that the correlations between length-weight and circumference-weight are high. The correlations of

circumference-rows of kernels, and length-circumference are considerable. The correlation of weight-rows of kernels is low, while that of length-rows is probably insignificant.

These correlations mean that the tendency towards a relative proportioning of length and circumference; and circumference and rows of kernels is considerably greater than that towards a relative proportioning of weight and rows of kernels or length and rows of kernels.

It seems somewhat disappointing that the correlation coefficients differ so widely, as this fact complicates the problem of assessing the influence of the selection of parents in a precise measure of heredity. The wide difference for different pairs of characters may be compared with the correlations between different pairs of characters of the human body, where it has been found that between measurement of the long bones of the arms and legs a high correlation exists, while between different measurement of the skull a much lower correlation exists.*

TABLE 1.—CORRELATION AMONG CERTAIN CHARACTERS OF EARS OF CORN 400 SERIES. CROP 1907. SEED: LOW PROTEIN BY MECHANICAL SELECTION

	Yield †	Value of r for length and circumference	Value of r for length and rows	Value of r for circumference and rows
401	68.4	0 423±0.019	−0.044±0.024	0.501±0.019
402	69.9	0.438±0.019	+0.007±0.024	0.446±0.020
403	69.4	0.312±0.020		0.484±0.018
404	75.9	0.462±0.018		0.548±0.017
405	66.6	0.452±0.018		0.502±0.019
406	84.6	0.403±0.019		0.470±0.018
407	68.6	0.282±0.021		0.440±0.020
408	84.1	0.278±0.021		0.554±0.016
409	71.4	0.203±0.021		0.487±0.019
410	95.6	0.411±0.017		0.432±0.018

		Value of r for length and weight	Value of r for rows and weight	Value of r for weight and circumference
401		0.781±0.008	0.275±0.023	0.768±0.009
405		0.786±0.009	0.223±0.024	0.721±0.011

*Biometrika, Vol. 1, pp. 408-467.

†Computed at 80 lb. per bushel of ear corn.

TABLE 2.—CORRELATION AMONG CERTAIN CHARACTERS OF EARS OF CORN 500 SERIES. CROP 1908. SEED: HIGH PROTEIN BY MECHANICAL SELECTION

	Yield *	Value of r for length and circumference	Value of r for length and rows	Value of r for circumference and rows
501	37.5	0.590±0.014	0.090±0.025	0.514±0.019
502	45.0	0.528±0.015	0.061±0.026	0.506±0.019
503	33.1	0.562±0.015	0.120±0.027	0.432±0.022
504	39.3	0.526±0.016		0.486±0.021
505	39.6	0.519±0.016		0.608±0.017
506	76.1	0.422±0.015		0.480±0.016
507	46.1	0.385±0.018		0.444±0.020
508	74.1	0.360±0.016		0.517±0.016
509	39.8	0.444±0.017		0.508±0.019
510	75.0	0.344±0.017		0.497±0.015

		Value of r for length and weight	Value of r for rows and weights	Value of r for weight and circumference
501		0.855±0.006	0.345±0.021	0.771±0.009
505		0.871±0.005	0.348±0.021	0.763±0.009

* Computed at 80 lb. per bushel of ear corn.

TABLE 3.—CORRELATION AMONG CERTAIN CHARACTERS OF EARS OF CORN 400 SERIES. CROP 1909. SEED: LOW PROTEIN BY MECHANICAL SELECTION

	Yield *	Value of r for length and circumference	Value of r for length and rows	Value of r for circumferences and rows
401	46.2	0.548±0.016	0.004±0.027	0.452±0.021
402	38.6	0.623±0.016	0.027±0.034	0.466±0.026
403	48.4	0.453±0.118	-0.044±0.026	0.514±0.022
404	43.6	0.534±0.016		0.463±0.022
405	47.2	0.461±0.018		0.487±0.022
406	57.6	0.506±0.017		0.482±0.019
407	46.0	0.443±0.019		0.524±0.020
408	58.8	0.432±0.018		0.428±0.021
409	51.6	0.409±0.019		0.458±0.022
410	72.6	0.539±0.012		0.425±0.018

TABLE 3.—*Concluded.*

		Value of r for length and weight	Value of r for weight and rows	Value of r for weight and circumference
401		0.818±0.008	0.216±0.025	0.840±0.007
402		0.844±0.008	0.225±0.032	0.746±0.012
403		0.815±0.008	0.178±0.027	0.648±0.013
404		0.801±0.010	0.212±0.029	0.757±0.012
405		0.810±0.008	0.229±0.027	0.728±0.011
406		0.791±0.008		
407		0.800±0.008		
408		0.798±0.008		
409		0.785±0.008		
410		0.843±0.005		

* Computed at 50 lb. per bushel.—Shelled corn (dry substance).

TABLE 4.—CORRELATION AMONG CERTAIN CHARACTERS OF EARS OF CORN ILLINOIS CORN. CROPS 1907, 1908, 1909

	Value of r for length and circumference	Value of r for circumference and rows	Value of r for length and rows	Value of r for length and weight
Crop 1907				
High protein	0.202±0.031	0.490±0.026	−0.051±0.032	0.822±0.013
Low proein	0.368±0.028	0.520±0.025	−0.017±0.034	0.725±0.016
High oil	0.317±0.029	0.431±0.027		0.838±0.009
Low oil	0.128±0.035	0.562±0.024		0.727±0.017
Crop 1908				
High protein	0.310±0.027	0.435±0.025	−0.106±0.030	0.776±0.012
Low protein	0.293±0.027	0.488±0.024	−0.017±0.031	0.809±0.012
High oil	0.132±0.027	0.464±0.022		0.673±0.015
Low oil	0.569±0.020	0.459±0.025		0.868±0.007
Crop 1909				
High protein	0.183±0.030	0.592±0.021	−0.081±0.035	0.681±0.016
Low protein	0.437±0.022	0.305±0.029		0.770±0.011
High oil	0.299±0.025	0.335±0.026		0.769±0.013
Low oil	0.255±0.024	0.480±0.020		0.675±0.014

APPENDIX ON THE MATHEMATICAL THEORY OF CORRELATION.

1. Mathematical function.—A variable y is said to be a mathematical *function* of a variable x if they are so related that to assigned values of x there correspond definite values of y. Thus, if $y=2x+4$, y is a function of x; since, for any assigned value of x, we can compute y. Those who are familiar with analytic geometry know that a curve is useful for representing and following the variations of a mathematical function.

We shall assume, in the present treatment of correlation, a knowledge* of the use of a system of co-ordinate axes to represent numbers and functions.

In order to place the notion of correlation on a precise basis, we lay down the following special

2. Definition.†—*Two measurable characters of an individual or of related individuals are said to be correlated if to a selected series of sizes of the one there correspond sizes of the other whose mean values are functions of the selected values.* The word "sizes" is used in the sense of numerical measure, and the function is to be different from zero for some of the selected values.

To be concrete, we may think, for example, of measuring the correlation between length and circumference of ears of corn, or the correlation of fathers and sons with respect to stature.

To render the above definition in symbolic language and to develop the methods of determining the function mentioned in the definition are the first points in the application of mathematics to the theory of correlation. For this purpose, let x and y be variables such that $y=f(x)$ gives the mean value of a system of variates which correspond to a selected x. Suppose the following system of corresponding values results from measurement: (x', y'), (x'', y''), . . . , $(x^{(n)}, y^{(n)})$, where n is a large number indicating the total number of pairs observed. These observations are said to form a total population or universe of observations. As it is more convenient to deal with the deviations of the observations

*See Davenport's Principles of Breeding, pp. 687, 689.

†*Philosophical Transactions of the Royal Society,* Vol. 187A, pp. 256-257.

from their mean values than with the observations themselves, let (x_1, y_1), (x_2, y_2), . . ., (x_n, y_n) represent the deviations of the observations from their mean values. These deviations may be conveniently represented with respect to co-ordinate axes (Fig. 5). The origin then represents the mean of the two characters. In fact, we may think of the co-ordinate axes as passing through the mean of the table and drawn parallel to arrays. The vertical parallel lines of the figure may then be looked upon as separating the observations into arrays. The values of the y's which correspond to a given class mark x are said to form a y-array. Suppose there are s such arrays.

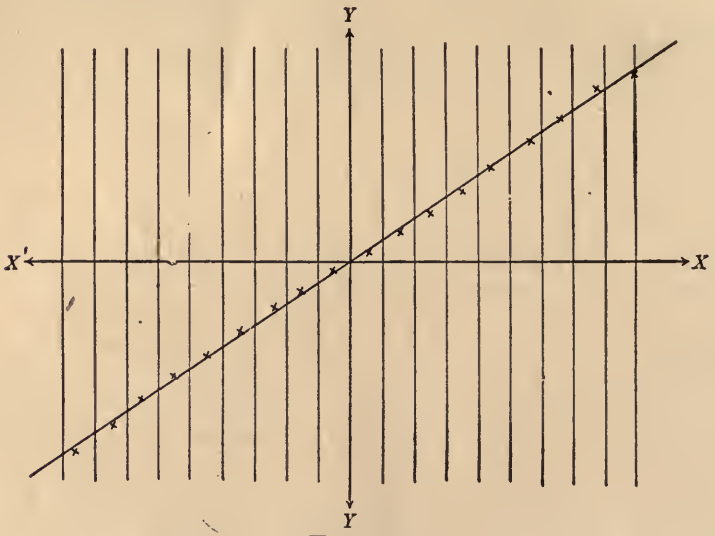

FIG. 5.

Let the crosses (×) in Fig. 5 represent the means of the *y's* in each of the *s*-arrays. If correlation exists, these means do not lie at random over the field, but arrange themselves more or less in the form of a smooth curve called the "curve of regression." This curve is a crude picture of the function which defines the correlation of the *y*-character relative to the *x*-character. Experience has shown that, in many sets of measurements, this line is approximately a straight line. For this reason, and for simplicity, the line subjected to the condition that the sum of the squares of the deviations (measured parallel to the *y*-axis and weighted with number of points in array) of the means from it shall be a minimum, is called the "line of regression." When the means lie exactly on the line, the regression is said to be "truly linear."

Let $y=mx+b$ be the function which represents the line of regression, then the problem of determining the line is that of determining m and b by means of the above minimal condition. The algebraic details of subjecting a line to this minimal condition are well known to those familiar with the method of least squares or the method of moments. The equation of the resulting line is

$$y = r\frac{\sigma_y}{\sigma_x}x\ , \text{- - - -}\ (1)$$

where σ_x is the standard deviation of the population with respect to the x-character, σ_y is the standard deviation with respect to the y-character, and r is the correlation coefficient given by

$$r = \frac{\Sigma x\,y}{n\sigma_x\,\sigma_y}\ ,$$

where the summation is extended to every pair of corresponding variates of the population. Similarly, the regression of the x character on the y character is given by

$$x = r\frac{\sigma_x}{\sigma_y}\,y \text{- - - -}(2)$$

It should be noted that (2) cannot be obtained by solving (1) for x, for the reason that the correspondence is one between selected values and means.

3. **Standard deviation of arrays.**—Suppose that regression is truly linear, so that the means of the y-arrays fall on the line $y = r\frac{\sigma_y}{\sigma_x}x$; and, for the present, assume that the standard deviations of arrays are equal. Then the standard deviation of an array is given by

$$\frac{\Sigma\left(y - r\frac{\sigma_y}{\sigma_x}x\right)^2}{n}\ ,$$

where the summation extends to the entire population.

$$\frac{\Sigma\left(y - r\frac{\sigma_y}{\sigma_x}x\right)^2}{n} = \frac{\Sigma y^2}{n} - \frac{2\sigma_y}{\sigma_x}\,r\,\frac{\Sigma x\,y}{n} + \frac{r\sigma_y^2}{\sigma_x^2}\,\frac{\Sigma x^2}{n}$$

$$= \sigma_y^2 - 2r^2\sigma_y^2 + r^2\sigma_y^2$$

$$= \sigma_y^2\left(1 - r^2\right) \text{- - - -}\ (3)$$

Hence, the standard deviation of a y-array is obtained from the standard deviation σ_y by multiplying σ_y by $\sqrt{1-r^2}$.

If the standard deviations of parallel arrays are unequal, then $\sigma_y\sqrt{1-r^2}$ is simply sort of an average value for the standard deviation of an array.

Since the first member of (3) is a sum of squares divided by n, the second member must be positive. Hence $-1 \leqq r \leqq 1$.

This proves that the correlation coefficient takes values not greater than +1 nor less than —1.

Equation (3) shows further that if r=+1 all the individual points plotted from observations must lie on the line of regression, and we can in this case, when one character is given, tell exactly the magnitude of the associated character. Further, the ratio

$$\frac{x_1}{y_1} = \frac{x_2}{y_2} = \text{ - - - } = \frac{x_n}{y_n} = \text{a positive constant.}$$

Similarly, if r=—1, the individual points plotted all lie on the line of regression, but

$$\frac{x_1}{y_1} = \frac{x_2}{y_2} = \text{ - - - } = \frac{x_n}{y_n} = \text{a negative constant.}$$

4. Correlations among three or more characters.—The theory of correlation can be extended to apply to any number of variables. However, the complexity of the algebraic expression for any number,—say n-variables—becomes so great that it does not seem well to present a more extended discussion here, except to say that the final result is expressed in standard deviations and correlation between systems of variates in sets of two, so that the problem is capable of reduction to the one which we have solved.

For the general case, the reader with considerable mathematical training is refered to the treatment by Karl Pearson in the Philosophical Transactions of the Royal Society, A, 187, 1896, and A, 200, 1903.

5. Correlation surfaces.—If our frequency distributions follow normal probability curves (Bulletin 119, pp. 30-31) there can be derived a surface

$$z = f(x, y)$$

such that f (x, y,) h. k gives, to within deviations due to random sampling, the number of the population with corresponding measurements in the region bounded by x=x, y=y, x=x+h, y=y+k, where the x and y are deviations from mean values and h and k are any small numbers. For a considerable range of statistical data, this surface takes the form

$$z = \frac{n}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} C^{-\frac{1}{2(1-r^2)}\left[\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}-2r\frac{xy}{\sigma_x\sigma_y}\right]},$$

where e is the base of natural logarithms and the other symbols have been defined above. The symbol π equals 3.1416, the ratio of circumference to a diameter in the circle.

As the only parameters in this surface (aside from the total number n) are the standard deviations and the correlation coefficient, we have in the standard deviations and its correlation coefficient a perfect description of a normally distributed population. This fact adds much to the significance of r as a measure of correlation.

6. **Formula for the correlation coefficient r which are better adapted to numerical calculation.**—In the first place, the calculation of the means and standard deviation of both systems of variates should be done by the shorter method presented on pp. 9-11, Bulletin 119.

It may be well to give that method here in a more symbolic form to prepare the way for the modified formula for r adapted to calculation.

Let G represent a guess at the mean M given by

$$M = \frac{f_1 v_1 + f_2 v_2 + \cdots + f_s v_s}{f_1 + f_2 + \cdots + f_s}, \quad \cdots (1)$$

where the class marks and f's are corresponding frequencies. Also let c be the correction to the guess G which gives M. That is

$$M = G + c,$$

$$c = M - G = \frac{f_1 v_1 + f_2 v_2 + \cdots + f_s v_s}{f_1 + f_2 + \cdots + f_s} - G.$$

$$c = \frac{f_1 (v_1 - G) + f_2 (v_2 - G) + \cdots + f_s (v_s - G)}{f_1 + f_2 + \cdots + f_s} \quad \cdots (2)$$

Formula (2) gives the practical method of finding the correction to be applied to the guess to get the mean.

Next, the standard deviation is given by

$$\sigma^2 = \frac{f_1 (v_1 - M)^2 + f_2 (v_2 - M)^2 + \cdots + f_s (v_s - M)^2}{f_1 + f_2 + \cdots + f_s},$$

$$= \frac{f_1 (v - G - c)^2 + f_2 (v_2 - G - c)^2 + \cdots + f_s (v - G - c)^2}{f_1 + f_2 + \cdots + f_s},$$

$$= \frac{\Sigma f_t (v_t - G)^2 - 2c \Sigma f_t (v_t - G) + c^2 \Sigma f_t}{\Sigma f_t} ,$$

$$= \frac{\Sigma f_t (v_t - G)^2 - 2c \Sigma f_t (v_t - M + c) + c^2 \Sigma f_t}{\Sigma f_t} ,$$

$$= \frac{\Sigma f_t (v_t - G)^2}{\Sigma f_t} - c^2 \quad \text{- - - - (3)}$$

Formula (3) gives the practical method of calculating the standard deviation.

The value of r is given by

$$r = \frac{\Sigma x.y}{n \sigma_x \sigma_y},$$

where x and y represent deviations from the means, and the summation extends to every pair of corresponding variates.

Let G_x and G_y represent class marks near the means of the systems of variates indicated by subscripts, and C_x, C_y corrections to these class marks which give the correct mean values so that

$$M_x = G_x + C_x ,$$

$$M_y = G_y + C_y .$$

Let x′, y′ be deviations from G_x and G_y which correspond to deviations x, y from the mean. Then

$$x = x' - C_x ,$$

$$y = y' - C_y .$$

$$r = \frac{\Sigma (x' - C_x)(y' - C_y)}{n \sigma_x \sigma_y} ,$$

$$= \frac{\Sigma x' y' - C_y \Sigma x' - C_x \Sigma y' + \Sigma C_x C_y}{n \sigma_x \sigma_y} ,$$

$$= \frac{\Sigma x' y' - C_y \Sigma (x + C_x) - C_x \Sigma (y + C_y) + \Sigma C_x C_y}{n \sigma_x \sigma_y} ,$$

$$= \left(\frac{\Sigma x' y'}{n} - C_x C_y \right) \frac{1}{\sigma_x \sigma_y} .$$

This is a formula whose computation is shown on pp. 297-301.

UNIVERSITY OF ILLINOIS

Agricultural Experiment Station

BULLETIN No. 148—ABSTRACT

ON THE MEASUREMENT OF CORRELATION WITH SPECIAL REFERENCE TO SOME CHARACTERS OF INDIAN CORN

By HENRY L. RIETZ and LOUIE H. SMITH

URBANA, ILLINOIS, NOVEMBER, 1910

The bulletin is a technical presentation of the methods of determining the correlation coefficient for associated characters, together with considerable tabular matter giving the correlations for various strains of Indian corn. It is the purpose of this abstract to present the leading thought of the bulletin devoid of technical terms, and, omitting all reference to methods of calculation, to discuss briefly the meaning of the correlation coefficient, present the data involved, and assist the non-mathematical reader to an understanding of its significance. Anyone desiring to pursue the subject farther, particularly as to methods of calculation of the correlation coefficient, can secure the complete text upon request to the Agricultural Experiment Station.

Any one, who has at all considered the matter, is conscious that there is correlation of some characters, both in animals and plants. That is to say, that the different characteristics that go to make up the individual animal or plant do not exist independently of one another, but on the contrary are more or less correlated, or bound together by such physiological bonds as compel them to move more or less with reference to each other.

Among those who have not studied the matter carefully, but rely merely upon personal impressions derived from unsystematic observation, it appears that a notion prevails pretty commonly that characters are either perfectly correlated or entirely uncorrelated, the conception being that the characters are either absolutely bound together or else they move with complete independence. The truth is, however, that characters are seldom perfectly correlated, just as they are seldom independent. The mathematician follows accurate methods in determining precisely what correlations exist between characters in large populations. He has no method of determining the bond between two or more characters from a single or even a few individuals. He deals only with large numbers, and by his methods, he is able to distinguish very clearly whether, in general, two characters tend to move together or in opposition to each other, and approximately to what extent. If they move together, correlation is said to be positive. If they move in opposite directions, the one tending to increase proportionately as the other decreases, the correlation is said to be negative.

The present bulletin treats the precise methods of measuring this correlation. It is measured by a single number called the correlation coefficient, denoted by r, which may take values from -1 to 1, depending on how fluctuations in the two characters take place. If, in general, the characters fluctuate together, say either above or below the type, the value of (r) lies between 0 and 1 depending upon how closely the characters are correlated. If, in

general, two corresponding characters fluctuate in opposite directions, the correlation is between 0 and -1. The values r =+1, and -1, indicate respectively perfect positive correlation, and perfect negative correlation, while indifference of the characters to each others fluctuations leads to the value r=0 when very large numbers are used.

The general reader is not concerned with the methods by which these values are obtained. He is concerned only with the results, which are significant and extremely valuable, and which with a little practice become easily apprehended by the non-mathematical reader. But a single further word of introduction is necessary, and that has reference to the so-called probable error, a decimal always following the correlation coefficient, and preceded by the + or — sign. This probable error has no reference to mistakes which might be made in computation. It has reference to the fact that any value which may be determined would probably have been different if a larger number of individuals had been involved. For example, if it is desired to ascertain what is the weight of mature draft horses of a given breed, it could be obtained approximately by weighing 100 such horses. It could be ascertained with greater accuracy by weighing 1000 such horses, but there is no absolutely accurate way of determining the actual weight until every draft horse of that age in the world has been weighed. The so called probable error of a result is a number that enables us to set limits within which we may reasonably expect the result to be found if we should use larger numbers in establishing a result. For a more complete statement of the meaning and applicability of the probable error, see Bulletin 119 of this station.

The bulletin treats the correlations among four characters of ears of corn—length, circumference, weight, and number of rows of kernels. The practical bearing of such information, as is contained in the results, lies in the facts, (1) that in the selection of parents for one character, we should know how this tends to change other characters; (2) that the problem of the correlation of characters and yield requires, for its solution, in case of a selected parentage, a knowledge of the correlation of the characters among themselves in the general population from which parents are selected; (3) that the problems of inheritance of these characters requires a knowledge of these correlations.

The following tables give the correlation coefficients and probable error for a large number of determinations that have been made at the Agricultural Experiment Station, and if the reader will take the pains to compare the different correlation coefficients,

TABLE 1.—CORRELATION AMONG CERTAIN CHARACTERS OF EARS OF CORN 400 SERIES. CROP 1907. SEED: LOW PROTEIN BY MECHANICAL SELECTION

	Yield *	Value of r for length and circumference	Value of r for length and rows	Value of r for circumference and rows
401	68.4	0.423±0.019	−0.044±0.024	0.501±0.019
402	69.9	0.438±0.019	+0.007±0.024	0.446±0.020
403	69.4	0.312±0.020		0.484±0.018
404	75.9	0.462±0.018		0.548±0.017
405	66.6	0.452±0.018		0.502±0.019
406	84.6	0.403±0.019		0.470±0.018
407	68.6	0.282±0.021		0.440±0.020
408	84.1	0.278±0.021		0.554±0.016
409	71.4	0.203±0.021		0.487±0.019
410	95.6	0.411±0.017		0.432±0.018

		Value of r for length and weight	Value of r for rows and weight	Value of r for weight and circumference
401		0.781±0.008	0.275±0.023	0.768±0 009
405		0.786±0.009	0.223±0.024	0.721±0.011

* Computed at 80 lb. per bushel of ear corn.

TABLE 2.—CORRELATION AMONG CERTAIN CHARACTERS OF EARS OF CORN 5 0 SERIES. CROP 1908. SEED: HIGH PROTEIN BY MECHANICAL SELECTION

	Yield *	Value of r for length and circumference	Value of r for length and rows	Value of r for circumference and rows
501	37.5	0.590±0.014	0.090±0.025	0.514±0.019
502	45.0	0.528±0.015	0.061±0.026	0.506±0.019
503	33.1	0.562±0.015	0.120±0.027	0.432±0.022
504	39.3	0.526±0.016		0.486±0.021
505	39.6	0.519±0.016		0.608±0.017
506	76.1	0.422±0.015		0.480±0.016
507	46.1	0.385±0.018		0.444±0.020
508	74.1	0.360±0.016		0.517±0.016
509	39.8	0.444±0.017		0.528±0.019
510	75.0	0.344±0.017		0.497±0 015

		Value of r for length and weight	Value of r for rows and weights	Value of r for weight and circumference
501		0.855±0.006	0.345±0.021	0.771±0.009
505		0.871±0.005	0.348±0.021	0.763±0.009

* Computed at 80 lb. per bushel of ear corn.

TABLE 3.—CORRELATION AMONG CERTAIN CHARACTERS OF EARS OF CORN 400 SERIES. CROP 1909. SEED: LOW PROTEIN BY MECHANICAL SELECTION

	Yield *	Value of r for length and circumference	Value of r for length and rows	Value of r for circumferences and rows
401	46.2	0.548±0.016	0.004±0.027	0.452±0.021
402	38.6	0.623±0.016	0.027±0.034	0.466±0.026
403	48.4	0.453±0.118	—0.044±0.026	0.514±0.022
404	43.6	0.534±0.016		0.463±0.022
405	47.2	0.461±0.018		0.487±0.022
406	57.6	0.506±0.017		0.482±0.019
407	46.0	0.443±0.019		0.524±0.020
408	58.8	0.432±0.018		0.428±0.021
409	51.6	0.409±0.019		0.458±0.022
414	72.6	0.539±0.012		0.425±0.018

		Value of r for length and weight	Value of r for weight and rows	Value of r for weight and circumference
401		0.818±0.008	0.216±0.025	0.840±0.007
402		0.844±0.008	0.225±0.032	0.746±0.012
403		0.815±0.008	0.178±0.027	0.648±0.013
404		0.801±0.010	0.212±0.029	0.757±0.012
405		0.810±0.008	0.229±0.027	0.728±0.011
406		0.791±0.008		
407		0.800±0.008		
008		0.798±0.008		
409		0.785±0.008		
410		0.843±0.005		

* Computed at 50 lb. per bushel. Shelled corn (dry substance).

he will learn something of the way corn behaves under a variety of conditions.

The material for this study is furnished by the crops obtained from a number of the regular experiment plots that are being conducted for different purposes. These plots belong to two different groups, one of which is devoted primarily to soil investigation (Tables 1-3), and the other to experiments in corn breeding (Table 4).

The soil plots are designated as the 400 and 500 series. They are devoted to a two year rotation consisting of corn alternating with oats, that is to say, corn occupies the 400 series one year and the 500 series the next year. Each series is divided into ten plots, and various soil treatments are applied. For a detailed account of the arrangement and the soil treatment of these plots, the reader is referred to the bulletin of which this is an abstract, or to Bulletin 125 of this station.

TABLE 4.—CORRELATION AMONG CERTAIN CHARACTERS OF EARS OF CORN ILLINOIS CORN. CROPS 1907, 1908, 1909

	Value of r for length and circumference	Value of r for circumference and rows	Value of r for length and rows	Value of r for length and weight
Crop 1907				
High protein .	0.202±0.031	0.490±0.026	—0.051±0.032	0.822±0.013
Low pro ein ..	0.368±0.028	0.520±0.025	—0.017±0.034	0.725±0.016
High oil	0.317±0.029	0.431±0.027		0.838±0.009
Low oil	0.128±0.035	0.562±0.024		0.727±0.017
Crop 1908				
High protein .	0.310±0.027	0.435±0.025	—0.106±0.030	0.776±0.012
Low protein ..	0.293±0.027	0.488±0.024	—0.017±0.031	0.809±0.012
High oil	0.132±0.027	0.464±0 022		0.673±0.015
Low oil	0.569±0.020	0.459±0.025		0.868±0.007
Crop 1909				
High protein .	0.183±0.030	0.592±0.021	—0.081±0.035	0.681±0.016
Low protein ..	0.437±0.022	0.305±0.029		0.770±0.011
High oil	0.299±0.025	0.335±0.026		0.769±0.013
Low oil	0.255±0.024	0.480±0.020		0.675±0.014

The variety of corn grown upon these plots has been two strains of Leaming which has been under selection for high protein and low protein content respectively. In 1907 and 1909 the seed corn planted was low in protein content while in 1908 it was high in protein.

The corn breeding plots from which the material for Table 4 was taken, represents four lines of selection that have been under way since 1896 for high protein content, low protein content, high oil content and low oil content.

With a set of four characters under consideration, there are six pairs of characters, between each pair of which the correlation can be determined. From the results of the tables, it will be observed that the correlations for some plots are given for each of these six pairs of characters. The tables include the results of 141 determinations of correlation, and should give a good general notion of the values of these correlations for the corn under consideration.

In Table 1 are given, correlations between length and circumference, length and number of rows, circumference and number of rows, length and weight, weight and rows of kernels, weight and circumference, for some plots of low protein corn, differently treated, crop of 1907. A careful study of this Table shows, first of all, a considerable tendency for length and circumference to move together, that is, for long ears to be large in circumference,

but that this correlation varies greatly in the different plots, ranging all the way from 0.203 to 0.462. Second, there is practically no correlation between the length of ear and the number of rows it contains, that is to say, one is no index whatever to the other. Third, there is a fairly high positive correlation between circumference and the number of rows, which means that the large ears have in general more rows than the small ears. However, the correlation in no case approaches very near to 1.0, which means that the large ears have not only more rows than the smaller ones, but the kernels are larger. There is a high correlation for length-weight, and weight-circumference, but a rather low correlation between weight and rows of kernels.

In Table 2 are also shown ten plots of high protein corn, differently fertilized, crop of 1908. The same general traits are maintained as in the former table, excepting that the correlation runs somewhat higher between length and circumference; and, that there is perhaps a slight positive correlation between the length of the ear and the number of rows that it contains.

Table 3 exhibits the correlation for the same series as shown in Table 1, but for the crop of 1909, and a more complete list of correlations between length-weight, weight-number of rows, and weight-rows of kernels. In these later determinations, we find what we should now expect, namely, a high correlation between length and weight and between weight and circumference, and a rather low correlation between weight and the number of rows.

Table 4 exhibits certain correlations for the so called Illinois corn, which, as stated above, consists of four strains bred for chemical composition. A gross comparison of this table will show that the correlation in any two characters varies in the same strain of corn in different years as it does by different methods of treatment. For example, the correlation between length and circumference in high protein corn, crop of 1907, was 0.202. The next year it was 0.310, and the next, 0.183. The reader will be interested in making this same sort of comparison for other strains of corn and for other characters.

Arranging the pairs of characters in descending order as to correlation, we have the following order:

(1) Length and weight.
(2) Circumference and weight.
(3) Circumference and rows of kernels.
(4) Length and circumference.
(5) Weight and rows of kernels.
(6) Length and rows of kernels.

For this arrangement, the odds are pretty large except in the case of (3) and (4), and possibly of (1) and (2).

As a sort of general conclusion, we may say that correlations for length-weight and circumference-weight are high. The correlation for circumference-rows of kernels and length-circumference are fairly high. The correlation of weight-rows of kernels is low, while that of length-rows of kernels is probably, in general, insignificant.